lapse of time can exist without an objective lapse of time, no reason can be given why an objective lapse of time should be assumed at all" (p. 206; see Yourgrau 1999).

Here, then, is another example of the Janus-faced quality of Gödel's thinking, presaged already in his arithmetization of metamathematics—contributing mathematically to "the left" while at the same time, as he sees it, pointing to "the right."

See also Gödel's Incompleteness Theorems; Logic, History of; Mathematics, Foundations of.

Bibliography

PRIMARY SOURCES

"Russell's Mathematical Logic" [1944]. In *Collected Works*, vol. 2. New York: Oxford University Press, 1990.

"Some Observations about the Relationship between Theory of Relativity and Kantian Philosophy" [1946/9]. In *Collected Works*, vol. 3. New York: Oxford University Press, 1994.

"What Is Cantor's Continuum Problem?" [1947; 1964]. In *Collected Works*, vol. 2. New York: Oxford University Press, 1990.

"Some Basic Theorems on the Foundations of Mathematics and Their Implications" [1951]. In *Collected Works*, vol. 3. New York: Oxford University Press, 1994.

"The Modern Development of the Foundations of Mathematics in the Light of Philosophy" [1961], transl. E. Köhler and H. Wang. In *Collected Works*, vol. 3. New York: Oxford University Press, 1994.

"Ontological Proof" [1970]. In *Collected Works*, vol. 3. New York: Oxford University Press, 1994.

"On an Extension of Finitary Mathematics Which Has Not Yet Been Used" Translated by L. F. Boron and revised by K. Gödel. In *Collected Works*, vol. 2. New York: Oxford University Press, 1990.

"A Remark about the Relationship between Relativity Theory and Idealistic Philosophy" In *Collected Works*, vol. 2. New York: Oxford University Press, 1990.

"Is Mathematics Syntax of Language?" In *Collected Works*, vol. 3. New York: Oxford University Press, 1994.

SECONDARY SOURCES

Buldt, B., et al., eds. *Kurt Gödel: Wahrheit und Beweisbarkeit*. Vienna: Hölder-Pichler-Tempsky, 2002.

Dawson, J. *Logical Dilemmas: The Life and Work of Kurt Gödel*. Wellesley: A. K. Peters, 1995.

Feferman, S. "Kurt Gödel: Conviction and Caution." *Philosophia Naturalis* 21 (1984): 546–562.

Frege, G. *Begriffsschrift: A Formula Language, Modeled upon That of Arithmetic, for Pure Thought* [1879]. Translated by S. Bauer-Mengelberg. In *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*, edited by J. van Heijenoort. Cambridge, MA: Harvard University Press, 1967.

Frege, G. *The Foundations of Arithmetic* [1884]. Translated by J. L. Austin. Evanston, IL: Northwestern University Press, 1980.

Kreisel, G. "Kurt Gödel." *Biographical Memoirs of Fellows of the Royal Society* 26 (1980): 149–224.

Parsons, C. "Platonism and Mathematical Intuition in Kurt Gödel's Thought." *The Bulletin of Symbolic Logic* 1 (1) (1995a): 44–74.

Parsons, C. "Quine and Gödel on Analyticity." In *On Quine: New Essays*, edited by P. Leonardi., Cambridge, MA: Cambridge University Press, 1995.

van Atten, M., and J. Kennedy. "On the Philosophical Development of Kurt Gödel." *The Bulletin of Symbolic Logic* 9 (4) (December 2003): 425–476.

Wang, H. *From Mathematics to Philosophy*. New York: MIT Press, 1974.

Wang, H. *A Logical Journey: From Gödel to Philosophy*. Cambridge, MA: MIT Press, 1996.

Yourgrau, P. *Gödel Meets Einstein: Time Travel in the Gödel Universe*. Chicago: Open Court, 1999.

Yourgrau, P. "Review Essay: Hao Wang, *Reflections on Kurt Gödel*." *Philosophy and Phenomenological Research* 1 (1989): 391–408.

Palle Yourgrau (1996, 2005)

GÖDEL'S INCOMPLETENESS THEOREMS

The axiomatic method is at the heart of mathematics. The work of mathematicians is to derive the consequences of axioms. According to Euclid, axioms are evidently true, and deduction from them is a powerful method of learning new truths. The rise of non-Euclidean geometry disrupted the carefree connection between truth and proof and led many modern thinkers to adopt the formalistic attitude that the mathematician's sole endeavor is to work out the consequences of axioms, taking no professional interest in inquiring what, if anything, the axioms are true of.

In 1931 Kurt Gödel proved a deep theorem that showed that deduction from axioms cannot be all there is to mathematical understanding. Gödel showed that, for whatever system of truths of number theory we choose to regard as axiomatic, there will be statements of basic arithmetic that we can recognize as true even though they are not consequences of the axioms. That there are truths not derivable from our axioms is hardly surprising; nobody ever promised us omniscience. What is surprising is that there are arithmetical statements *we can recognize as true* even though they are not derivable, so that no system of axioms we can write down fully captures our arithmetical understanding. Moreover this situation holds not only for systems of axioms we are capable of producing today but also for whatever systems we may devise in the future.

Gödel's true, unprovable sentence is obtained by using strings of numbers to encode strings of symbols, thereby reducing statements about language to statements about numbers. Under such a coding Gödel's sentence says that the system of axioms is consistent. Of course if we accept the axioms, we regard the axioms as true, so we certainly regard them as consistent. But even though adopting the axioms means accepting their consistency, the statement that the axioms are consistent cannot be proved from the axioms. We could adopt the thesis that the axioms are consistent as a new axiom. This would give us a new, larger system of axioms that can prove the consistency of the old system but not the consistency of the new system. We can continue the process of adding consistency statements repeatedly, but however far we go we shall never catch up with Gödel. No consistent system that includes basic arithmetic can prove its own consistency.

Gödel's result has important corollaries, notably, Church's theorem (1936) that there is no algorithm for testing whether a sentence is logically valid and Tarski's theorem (1935) that the set of true sentences of a language cannot be defined within the language itself.

THE LANGUAGE OF ARITHMETIC

Gödel's results apply to the language of arithmetic, which is an artificial language for formalizing reasoning about the natural numbers, and to other languages into which the language of arithmetic can be translated. To state his results we need to specify the language exactly. As numerals, the language uses "0" and expressions obtained from "0" by repeatedly prefixing "*S*," which stands for the successor function. The numeral for 3 is "SSS0," which we abbreviate "$\underline{3}$." The language also contains function signs "+" "×" and "E," for addition, multiplication, and exponentiation, so that the *terms* of the language make up the smallest class that contains the numeral "0" and the variables v_0, v_1, v_2, v_3, …, and that contains $S\tau$, $(\tau+\rho)$, $(\tau\times\rho)$, and $(\tau E\rho)$ whenever it contains τ and ρ. In the exposition here we shall sometimes use other letters as variables in place of the official v_is, so as to reduce the proliferation of subscripts. Including "E" as a primitive operation is not strictly necessary, as we shall see below, but it enables us to get off to a fast start.

A term without variables is *closed*. Rules that we learned in elementary school enable us to calculate the numerical value of each closed term. A term with *n* variables represents an *n*-ary function, calculable by a grade-school algorithm.

The *atomic formulas* take the form $\tau = \rho$ or $\tau \leq \rho$, where τ and ρ are terms, and the *formulas* constitute the smallest class containing the atomic formulas and containing $\sim \phi$, $(\phi \vee \psi)$, and $(\exists v_i)\phi$, whenever it contains ϕ and ψ. An occurrence within a formula of the variable v_i is *bound* if it occurs within some subformula that begins with $(\exists v_i)$, and it is *free* otherwise. A formula without free variables is a *sentence*; it is sentences that are either true or false. The symbols for conjunction ("$\wedge$"), the conditional ("$\rightarrow$"), the biconditional ("$\leftrightarrow$"), universal quantification ("$(\forall v_i)$"), and the less-than relation ("$<$") are treated as defined.

Where v_i does not occur within the term τ, we use $(\exists v_i{\leq}\tau)\phi$ and $(\forall v_i{\leq}\tau)\phi$ to abbreviate $(\exists v_i)(v_i{\leq}\tau \wedge \phi)$ and $(\forall v_i)(v_i{\leq}\tau \rightarrow \phi)$.These are *bounded quantifiers*, and a formula with no quantifiers that are not bounded is a *bounded formula*. For example 'v_0 is prime' is formalized by the bounded formula '$(SS0{\leq}v_0 \wedge (\forall v_1{\leq}v_0)(\forall v_2{\leq}v_0)(v_0 = (v_1 \times v_2) \rightarrow (v_1 = S0 \vee v_2 = S0)))$'. A set or relation is said to be *bounded* if it is the extension of a bounded formula.

We can test whether an atomic sentence is true by grade-school algorithms; "true," that is, in the standard model consisting of the natural numbers 0,1,2,3, … Any bounded sentence is demonstrably equivalent to a truth-functional combination of atomic sentences, since bounded quantifiers can be cashed out as long but finite disjunctions and conjunctions. Thus we have an algorithm for determining the truth value of a bounded sentence. It follows that every bounded set or relation is decidable; that is, there is an algorithm for testing membership in the set or relation. If *S* is the extension of the bounded formula $\sigma(x_0)$, we can test whether $n \in S$ by asking whether $\sigma(\underline{n})$ is true.

The Σ *formulas* are obtained by prefixing a block of existential quantifiers to a bounded formula, and their extensions are *recursively enumerable* sets and relations. Any recursively enumerable set is the extension of a formula obtained by prefixing a single existential quantifier to a bounded formula, since $(\exists x_1)(\exists x_2)\ldots(\exists x_n)\phi$ is equivalent to $(\exists x_0)(\exists x_1{\leq}x_0)(\exists x_1{\leq}x_0)\ldots(\exists x_n{\leq}x_0)\phi$. (The same goes for recursively enumerable relations; in the future we shall let this go without comment.) The union and intersection of recursively enumerable sets are recursively enumerable, since $((\exists y)\phi(x,y) \vee (\exists z)\psi(x,z))$ and $((\exists y)\phi(x,y) \wedge (\exists z)\psi(x,z))$ are, respectively, logically equivalent to $(\exists y)(\exists z)(\phi(x,y) \vee (\psi(x,z))$ and $(\exists y)(\exists z)(\phi(x,y) \wedge (\psi(x,z))$ (assuming bound variables have been chosen so as to avoid conflicts). If $\chi(x,y,z)$ is bounded and τ is a term, $\{x: (\exists y{\leq}\tau)(\exists z)\chi(x,y,z)\}$ and $\{x:(\forall y{\leq}\tau)(\exists z)\chi(x,y,z)\}$ are both recursively enumerable since they are the extensions of $(\exists z)(\exists y{\leq}\tau)\chi(x,y,z)$ and $(\exists w)(\forall y{\leq}\tau)(\exists z{\leq}w)\chi(x,y,z)$, re-

spectively. If a Σ sentence is true, we can show it is true by providing an appropriate witness.

Σ FORMULAS AND DECIDABILITY

A set of numbers is *effectively enumerable* if there is a mechanical procedure for listing the set, so that every member of the set turns up on the list eventually and nothing appears on the list that is not in the set. Every recursively enumerable set is effectively enumerable. To see this, we introduce the pairing function. $Pair(x,y) = \frac{1}{2}(x^2 + 2xy + y^2 + 3x + y)$ is a one-one correspondence between $\mathbb{N} \times \mathbb{N}$ and $\mathbb{N}$ (where $\mathbb{N}$ is the set of natural numbers). Define the functions 1st and 2nd so that $Pair(1st(z),2nd(z)) = z$. Given a recursively enumerable function $S = \{x_0: (\exists x_1)\sigma(x_0,x_1)\}$, with σ bounded, we can list S by the following algorithm: At stage n, test whether the sentence $\sigma(\underline{1st(n)}, \underline{2nd(n)})$ is true; if it is, add $1st(n)$ to the list.

Every set that is known to be effectively enumerable is recursively enumerable. This striking fact, together with a large body of evidence obtained by examining idealized models of computation and examining structural properties of effectively enumerable and recursively enumerable sets, has led to the general acceptance of the *Church-Turing thesis*: A set of natural numbers is effectively enumerable if and only if it is recursively enumerable.

A set of natural numbers is *decidable* if and only if there is an algorithm for testing membership in the set. A set can be effectively enumerable without being decidable, since, if we have a procedure for listing an infinite set, there will be no stage at which, from the fact that a given number has not yet turned up on the list, we can conclude that the number will never appear on the list. On the other hand if a set and its complement are both effectively enumerable then the set is decidable, and conversely. Defining a set to be *recursive* if it and its complement are both recursively enumerable, the Church-Turing thesis tells us that a set is decidable if and only if it is recursive.

An *unary partial function* is a set of ordered pairs f with the property that, whenever $<i,j>$ and $<i,k>$ are both in f, we have $j = k$. If $<i,j> \in f$, for some j, we say that i is in the *domain* of f, and we write $f(i) = j$. (Partial functions of more than one variable are defined similarly.) f is said to be *calculable* if there is an algorithm that, for given input i, gives the output $f(i)$ if i is in the domain of f, and yields no output at all if i is outside the domain of f. A unary partial function is calculable if and only if, *qua* binary relation, it is effectively enumerable. It follows according to the Church-Turing thesis that f is calculable if and only if it is recursively enumerable. If so, f is said to be a *partial recursive function*. (The notation is confusing—a collection of ordered pairs can be a partial recursive function without being a recursive relation—but entrenched.) A *total recursive function*—a partial recursive function whose domain is all of $\mathbb{N}$—will be a recursive relation, since if f is $\{<i,j>: (\exists x)\theta(\underline{i},\underline{j},x)\}$, with θ bounded, the complement of f is $\{<i,j>:(\exists x)(\exists y)(\sim y = \underline{j} \wedge \theta(\underline{i},y,x))\}$.

ARITHMETIZATION OF METAMATHEMATICS

The set-theoretic paradoxes, particularly Russell's paradox, had on David Hilbert much the same effect that Zeno's paradoxes had on Aristotle. Both thinkers came to realize that the idea of the infinite held great intellectual peril with the risk of contradiction at every turn. Unlike Aristotle, however, Hilbert was unwilling to banish the actual infinite from mathematical reasoning. Instead he proposed to develop the theory of infinite sets in such a way that we could be assured that no contradiction would ensue, by treating mathematical proofs as the objects of mathematical study, in the same way that earlier mathematicians had treated curves, planes, and numbers as objects of mathematical study. A mathematical proof is, after all, a finite object, even if the sentences that appear in the proof talk about infinite objects, and Hilbert proposed that a new science of *metamathematics* could show by finite means that set theory was free of contradiction, by showing that there is no finite path that leads from the axioms to "~ 0=0."

The great breakthrough in metamathematics was Gödel's proof, which showed that it was not necessary to go outside set theory or even outside arithmetic to carry out metamathematical investigations. By assigning numerical codes to formulas and finite strings of formulas, and by reducing properties of proofs to properties of their code numbers, it was possible to develop proof theory as a branch of number theory. This technique led to a great flowering of metamathematics even though as we shall see, it derailed Hilbert's plan.

The arithmetization of metamathematics proceeded in two stages. In the first stage numerical codes are assigned to simple symbols more-or-less arbitrarily, so that a formula, which is a string of simple symbols, can be coded as a sequence of numbers. Second we devise a method for encoding a finite sequence of numbers as a single number. This enables us to encode a formula as a single number. In this way a proof, which is a sequence of formulas, is encoded as a sequence of numbers, which is, in turn, coded as a single number.

We attack the second stage first. We already know how to use the function *Pair* to code a pair of natural numbers by a single number. We can encode a finite set of natural numbers by a single number by setting the code number of the finite set F, $Code(F)$, equal to $\sum_{i \in F} (2^i)$. *Code* provides a one-one correspondence between the set of finite sets of natural numbers and $\mathbb{N}$. The number n is the image under *Code* of the set of places in the binary decimal expansion of n in which "1"s appear. Finally, we encode the finite sequence $<k_0,k_1, \ldots ,k_m>$ as the number $Code(\{Pair(0,k_0), Pair(1,k_1), \ldots , Pair(m,k_m)\})$. Here we shall use an expression like "<3,2,1>" ambiguously to denote a sequence of length three and to denote the code number for that sequence, which is 448.

The relation that holds between k and n if k is an element of the set coded by n is defined by a bounded formula; abusing notation, we write "$\underline{k} \in \underline{n}$" to represent the statement that $(\exists i < (\underline{2}\mathrm{E}\underline{k}))(\exists j < \underline{n})\underline{n} = (i + ((\underline{2}\ \mathrm{E}\ \underline{k}) + (j \times (\underline{2}\mathrm{E}(\mathrm{S}\underline{k})))))$. The set of all code numbers of finite sequences is the extension of a bounded formula, as are the concatenation operation and the partial function that takes i and n to the ith member of the sequence coded by n (provided n codes a sequence of i or more elements). The simplicity of this technique for encoding a finite sequence of numbers by a single number is the motive for including exponentiation as a primitive operation.

The details of the assignment of numerical codes to terms and formulas are highly arbitrary. A motive for the particular choices here is to avoid fretting over parentheses. With each term τ, we associate a number $\ulcorner\tau\urcorner$, as follows: The numeral "0" is assigned <0,0>, and the variable x_i is assigned $<1,i>$. $\ulcorner S\tau\urcorner$ is $<2,\ulcorner\tau\urcorner>$, and $\ulcorner(\tau+\rho)\urcorner$, $\ulcorner(\tau\times\rho)\urcorner$, and $\ulcorner(\tau \mathrm{E} \rho z)\urcorner$ are $<3,\ulcorner\tau\urcorner,\ulcorner\rho\urcorner>$, $<4,\ulcorner\tau\urcorner,\ulcorner\rho\urcorner>$, and $<5,\ulcorner\tau\urcorner,\ulcorner\rho\urcorner>$, respectively.

A number x is a the code of a term just in case it is an element of a finite set s with the following property: For any element y of s, either $y = <0,0>$; or $y = <1,i>$, for some $i \leq y$; or $y = <2,z>$, for some z in s; or y is equal to one of $<3,z,w>$, $<4,z,\mathrm{w}>$, and $<5,z,\mathrm{w}>$, for some z and w in s. s represents a finite tree, with each node labeled by the code of a term, so that when a node is labeled by a complex term, nodes beneath it are labeled by the term's constituents and so that each leaf of the tree is labeled either by the code of "0" or by the code for a variable. This characterization is naturally written out as a Σ formula, showing that the set of (code numbers of) terms is recursively enumerable.

The set of terms is, in fact, recursive. To see this, we note that, if x is not a term, then the attempt to construct a labeled tree with x at its trunk winds up with at least one branch that does not terminate in either $\ulcorner 0\urcorner$ or a variable. More precisely, x does not encode a term if and only if there is a sequence $<x_0,x_1,\ldots,x_n>$ of numbers $\leq x$ with the following properties:

$x_0 = x$.

If x_i has the form $<2,y>$, then $i < n$ and $x_{i+1} = y$.

If x_i has one of the forms $<3,y,z>$, $<4,y,z>$, or $<5,y,z>$, then $i < n$ and either $x_{i+1} = y$ or $x_{i+1} = z$.

If $i < n$, x_i has one of the forms $<2,y>$, $<3,y,z>$, $<4,y,z>$, or $<5,y,z>$.

x_n does not have either of the forms $<0,0>$ or $<1,k>$.

This can readily be written out as a Σ formula, showing that the complement of the set of terms is recursively enumerable.

The function Z that takes a number n to the code number for the numeral $\underline{n}$ can be described by a recursive definition:

$$Z(0) = <0,0> = 5.$$
$$Z(m+1) = <2,Z(m)> = 8 + (2\mathrm{E}(Pair(1,Z(m)))).$$

We can convert this recursive definition into an explicit definition, using a quite general technique that Gödel obtained by refining an idea from Gottlob Frege's *Begriffschrift*. $Z(n) = k$ if and only if there is a sequence $<x_0,x_1,\ldots,x_n>$ with the following features:

$x_0 = <0,0>$.

For $m < n$, $x_{m+1} = <2,x_m>$.

$x_n = k$.

This characterization shows that Z is a total recursive function.

The function that associates a code $\ulcorner\phi\urcorner$ with each formula ϕ is again highly arbitrary. For τ and ρ terms, we let $<6,\ulcorner\tau\urcorner,\ulcorner\rho\urcorner>$ and $<7,\ulcorner\tau\urcorner,\ulcorner\rho\urcorner>$ be the codes of $\tau = \rho$ and $\tau \leq \rho$. For ϕ and ψ formulas, we let $<8,\ulcorner\phi\urcorner>$ be $\ulcorner\sim \phi\urcorner$, $<9,\ulcorner\phi\urcorner,\ulcorner\psi\urcorner>$ be $\ulcorner(\phi \vee \psi)\urcorner$, and $<10,i,\ulcorner\phi\urcorner>$ be $\ulcorner(\exists v_i)\psi\urcorner$. The proof that the set of codes of formulas is recursive is just like the corresponding argument for terms.

It is straightforward if somewhat laborious to verify, just by writing down an appropriate formula, that, for example, the arithmetical operations corresponding to forming the disjunction and the conjunction of two formulas, to prefixing a quantifier to a formula, and to substituting a given term for free occurrences of a variable in a formula are partial recursive functions. Also, for example, that the set of terms in which the variable v_{17} appears

and the set of formulas in which v_{123} appears free are recursive sets.

PROOFS AND COMPUTATIONS

Euclid's *Elements* deduces highly sophisticated geometric theorems as consequences of simple, intuitively obvious axioms. Aristotle, the father of logic, investigated the methods by which consequences are derived from axioms, identifying simple patterns of valid reasoning like the following so-called *syllogism*: "All men are animals. No stone is an animal. Therefore, no stone is a man." The methods of reasoning Euclid actually employed were far more sophisticated than the mere production of chains of syllogisms, however, and the ancients were generally content to take it as obvious that Euclid's deductions were legitimate, without demanding a detailed survey of deductive methods.

Meticulous nineteenth-century investigations revealed the surprising fact that, despite having been accepted by generations of scholars as the exemplar of deductive rigor, Euclid's proofs were often invalid. In proving a theorem he sometimes imported information from the accompanying diagram that was not justified by either the hypotheses of the theorem or the axioms. These investigations led to a search for fully precise methods of deduction in which one could have complete confidence. This search culminated in the widespread acceptance of a system of precise rules for the first-order predicate calculus—the logic governing the operators "$\vee$," "$\sim$," ($\exists v_i$), and "="—within which the deductions of classical mathematics can be formalized with scrupulous rigor.

With these rules in hand, we can capture the notion of logical consequence precisely, by pressing it from below and from above. It is clear that, if a sentence ϕ is a logical consequence of a theory (set of sentences) Γ, then it cannot be possible to choose a domain of discourse and semantic values for the nonlogical terms so as to make the members of Γ all true and ϕ false. Thus a necessary condition for a ϕ to be a logical consequence of Γ is that ϕ be true in every model of Γ. It is also clear, from examining the rules (for whichever of the standard textbook systems is convenient), that if ϕ derivable from Γ, ϕ is a logical consequence of Γ; this gives us a sufficient condition for logical consequence. Gödel's 1930 *Completeness Theorem* shows that these two conditions meet, so that if ϕ is true in every model of Γ then ϕ is derivable from Γ.

The Completeness Theorem applies equally well to any of many different logical calculi for first-order predicate logic. W.V. Quine developed a particularly convenient system with the following two properties: The (codes of the) axioms of logic form a recursive set; and each logical consequence of a theory Γ can be found at the end of a sequence of sentences, each member of which is either an axiom of logic, an element of Γ, or obtained from earlier members of the sequence by *modus ponens*, the rule that permits the deduction of ψ from ϕ and $(\phi \rightarrow \psi)$. (Such a sequence is a *proof* of the sentence from Γ.) Quine's axioms will not be written out here.

If Γ is recursive, the set of pairs $<s, \ulcorner\phi\urcorner>$ such that s is a proof of ϕ from Γ is a recursive relation, represented by a Σ formula we shall abbreviate "$\underline{s}\ B_\Gamma\ \underline{\ulcorner\phi\urcorner}$." (In terminology introduced below, "B_Γ" "binumerates" the relation.) We write "$Bew_\Gamma(\underline{\ulcorner\phi\urcorner})$" to abbreviate "$(\exists s)s\ B_\Gamma \underline{\ulcorner\phi\urcorner}$." Since "$Bew_\Gamma$," is Σ, the set of logical consequences of Γ is recursively enumerable.

William Craig noted a converse result: If the set of consequences of the theory Γ is recursively enumerable then Γ has the same consequences as some recursive set to axioms; Γ is, as they say, *recursively axiomatizable*. To see this, note that there is a bounded formula $\psi(x,y)$ such that the consequences of Γ constitute the set of sentences whose code numbers satisfy $(\exists y)\psi(x,y)$. Let Γ_{Craig} be the set of all sentences of the form $(\underline{m} = \underline{m} \wedge \theta)$, for which the pair $<\ulcorner\theta\urcorner, m>$ satisfies $\psi(x,y)$. Then Γ_{Craig} is recursive (bounded, in fact), and Γ_{Craig} and Γ are logically equivalent.

We would now like to see how any numerical computation by algorithm can be simulated by a logical deduction from basic arithmetical axioms. Q_E, a variant of *Robinson's arithmetic*, is the conjunction of the following nine statements:

$(\forall x)(x = 0 \leftrightarrow \sim (\exists y)x = Sy).$

$(\forall x)(\forall y)(Sx = Sy \rightarrow x = y)$

$(\forall x)(x + 0) = x.$

$(\forall x)(\forall y)(x + Sy) = S(x + y).$

$(\forall x)(x \times 0) = 0.$

$(\forall x)(\forall y)(x \times Sy) = ((x \times y) + x)$

$(\forall x)(x\ E\ 0) = S0.$

$(\forall x)(\forall y)(x\ E\ Sy) = ((x\ E\ y) \times x)$

$(\forall x)(\forall y)(x \leq y \leftrightarrow (\exists z)(x + z) = y).$

Q, which we shall talk about later on, is obtained from Q_E by deleting the two clauses involving exponentiation.

A straightforward induction on the complexity of terms shows that, for every closed term τ, there is a number m such that the sentence $\tau = \underline{m}$ is a theorem of Q_E.

Another induction shows that every bounded sentence is decidable (either provable or refutable) in Q_E. Since every true bounded sentence is provable in Q_E, it follows that every true Σ sentence is provable in Q_E, since we can prove an existential sentence by providing a witness. If S is a recursively enumerable set, it is the extension of some Σ formula σ. Because every true Σ sentence is provable in Q_E and (because Q_E is true) no false Σ sentence is provable, we have (where "$\vdash$" is provability):

For any n, $n \in S$ if and only if $Q_E \vdash \sigma(\underline{n})$.

We shall say that σ *enumerates* S in Q_E. (The same observation holds for recursively enumerable relations.)

We shall say a formula ϕ *binumerates* a set S in Q_E if and only if, for each n, we have:

$n \in S$ if and only if $Q_E \vdash \phi(\underline{n})$.

$n \notin S$ if and only if $Q_E \vdash \sim \phi(\underline{n})$.

If S is recursive then there is a bounded formula $\chi(x,y)$ such that $(\exists y)\chi(x,y)$ enumerates S in Q_E, and there is a bounded formula $\theta(x,y)$ such that $(\exists y)\theta(x,y)$ enumerates the complement of S in Q_E. To show that S is binumerable in Q_E, we need to show that S is enumerable by a formula whose negation enumerates the complement of S. Developing an idea of J. Barclay Rosser, Tarski, Mostowski, and Robinson showed that the following Σ formula does the job:

$$(\exists y)(\chi(x,y) \wedge \sim (\exists z < y)\theta(x,z)).$$

Clearly if ϕ binumerates S in Q_E, it binumerates S in any consistent theory that entails Q_E.

A formula $\psi(x,y)$ *functionally represents* a total function f in a theory if and only if, for each k, the following sentence is a consequence of the theory:

$$(\forall y)(\psi(\underline{k},y) \leftrightarrow y = f(\underline{k})).$$

If f is a total recursive function, we know that there is a formula $\phi(x,y)$ that binumerates f in Q_E. Tarski, Mostowski, and Robinson showed that the following formula functionally represents f in Q_E (and hence in any theory that entails Q_E):

$$(\phi(x,y) \wedge (\forall z < y)\sim \phi(x,z)).$$

THE FIRST INCOMPLETENESS THEOREM

We are now ready to see how to construct, for any recursively axiomatizable, true theory that includes Q_E, a true sentence that is not a consequence of the theory. The key to the construction is to see how to produce sentences that can talk about themselves so that we can construct a sentence that asserts its own unprovability. Such a sentence cannot be provable since if it were provable it would be a false consequence of the axioms. So the sentence must be true. To carry out this plan we use the following result, one of the masterpieces of modern mathematics:

GÖDEL'S SELF-REFERENCE LEMMA. For any formula $\psi(y)$, one can construct a sentence ϕ such that $Q_E \vdash (\phi \leftrightarrow \psi(\ulcorner\underline{\phi}\urcorner))$.

The hard part, the part that requires true genius, is to figure out what sentence to write down. The easy part is to verify that the sentence works. Here we shall only attempt the easy part.

Define a function f as follows: If m is the code of a formula $\chi(x,y)$ with only "x" and "y" free, let $f(m)$ be the code of the formula

$$(\exists x)(\exists y)((x = \underline{m} \wedge \chi(x,y)) \wedge \psi(y)).$$

Otherwise, $f(m) = 0$.

This definition can easily be written as a Σ formula, showing that f is a total recursive function. Consequently, there is a formula $\theta(x,y)$ that functionally represents f in Q_E. Let m be $\ulcorner(\theta(x,y)\urcorner$, and ϕ be the following sentence:

$$(\exists x)(\exists y)((x = \underline{m} \wedge \theta(x,y)) \wedge \psi(y)).$$

Then $\ulcorner\phi\urcorner = f(m)$, and so the following sentences are consequences of Q_E:

$(\forall y)(\theta(\underline{m},y) \leftrightarrow y = \ulcorner\underline{\phi}\urcorner)$.

$((\exists x)(\exists y)((x = \underline{m} \wedge \theta(x,y)) \wedge \psi(y)) \leftrightarrow \psi(\ulcorner\underline{\phi}\urcorner))$.

$(\phi \leftrightarrow \psi(\ulcorner\underline{\phi}\urcorner))$.

Let Γ be a consistent, recursive set of sentences that entails Q_E. Using the Self-reference Lemma, we can find a sentence γ so that $(\gamma \leftrightarrow \sim Bew_\Gamma(\ulcorner\underline{\gamma}\urcorner))$ is a consequence of Q_E; γ is called the *Gödel sentence* for Γ. If γ were a consequence of Γ, $\sim Bew_\Gamma(\ulcorner\underline{\gamma}\urcorner)$ would be a consequence of Γ, and also $Bew_\Gamma(\ulcorner\gamma\urcorner)$ would be a true Σ sentence, hence a consequence of Q_E, hence a consequence of Γ. This contradicts the consistency of Γ. So γ is unprovable, so that $Bew_\Gamma(\ulcorner\underline{\gamma}\urcorner)$ is false, and γ is true. Thus γ is our example of a true, unprovable sentences.

If Γ is true then Γ does not prove $\sim\gamma$ because $\sim\gamma$ is false, so that γ is undecidable in Γ. Let us say that a theory Δ is *ω-inconsistent* if there is a formula $\chi(x)$ such that $(\exists x)\chi(x)$ is a consequence of Δ, and yet, for each n, $\sim \chi(\underline{n})$ is a consequence of Δ. Every ω-consistent theory is consistent, so if Δ is a recursive, ω-consistent theory that entails Q_E, the Gödel sentence γ for Δ is a true sentence not provable in Δ. Hence, for each m, the sentence

$\sim \underline{m}\ B_\Delta \ulcorner\underline{\gamma}\urcorner$ is true, hence provable in Q_E, hence provable in Δ. It follows by ω-consistency that $Bew_\Delta \ulcorner\underline{\gamma}\urcorner$ is not a consequence of Δ, and so $\sim\gamma$ is not a consequence of Δ. Thus the assumption of ω-consistency, rather than truth, is enough to ensure that γ is undecidable in Δ. Because γ is unprovable in Δ, $\Delta \cup \{\sim\gamma\}$ is consistent, although ω-inconsistent. So consistency does not imply ω-consistency.

Gödel used γ to show that every ω-consistent, recursively axiomatizable theory that entails Q_E is incomplete, that is, that there are sentences that the theory cannot decide; this is the *First Incompleteness Theorem*. Rosser went a step farther, showing that the assumption of ω-consistency can be weakened to consistency. Rather than examine Rosser's proof, we shall derive his conclusion from a stronger result, one due, in essentials, to Tarski, Mostowski, and Robinson:

RECURSIVE INSEPARABILITY THEOREM. There is no recursive set that includes the consequences of Q_E and excludes all the sentences refutable in Q_E.

Suppose *C* were such a recursive set, and take a formula $\mu(x)$ that binumerates *C* in Q_E. The Self-reference Lemma gives a sentence ν such that $(\nu \leftrightarrow \sim \mu(\ulcorner\underline{\nu}\urcorner))$ is a consequence of Q_E. We derive a contradiction by examining two cases:

Case 1. $\nu \in C$. Then $Q_E \vdash \mu(\ulcorner\underline{\nu}\urcorner)$, and so $Q_E \vdash \sim \nu$. Thus υ is a sentence refutable in Q_E, and so it is excluded from C. Contradiction.

Case 2. $\upsilon \notin C$. Then $Q_E \vdash \sim \mu(\ulcorner\underline{\nu}\urcorner)$, and so $Q_E \vdash \nu$. Thus υ is a consequence of Q_E, and so an element of *C*. Contradiction.

Corollary. No consistent theory that entails Q_E has a recursive set of consequences.

This follows from the fact that, if a consistent theory entails Q_E, it excludes the sentences refutable in Q_E.

***Corollary* (*Rosser's Theorem*).** No consistent, recursively axiomatized theory that entails Q_E is complete.

If Γ is consistent, recursively axiomatized, and complete, then the complement of Γ is recursively enumerable, since it is the union of the set of non-sentences with the set of sentences whose negations are provable in Γ.

Corollary. No theory consistent with Q_E has a recursive set of consequences.

If Δ were such a theory then the set of sentences ψ such that $(Q_E \rightarrow \psi)$ is a consequence of Δ would be a consistent, recursive set of sentences, closed under consequence, that included Q_E.

***Corollary* (*Church's Theorem*).** The set of logically valid sentences in not recursive.

The valid sentences are the consequences of the empty theory, which is consistent with Q_E.

MATHEMATICAL INDUCTION

Q_E is a weak axiom system. It cannot prove the associative law of addition or multiplication, nor can it prove the commutative law of addition or multiplication. The system is weak because it leaves out the essential feature of the natural number system, the principle of mathematical induction, according to which any collection of natural numbers that includes 0 and is closed under the successor operation has to include all the natural numbers. *Modulo* Q_E, the principle is equivalent to the thesis that the natural numbers are well-founded, that is, that any nonempty collection of natural numbers has a least element.

Richard Dedekind showed that the system one gets from Q_E by adding the principle of mathematical induction completely characterizes the natural numbers. The system is *categorical*, that is, there is an isomorphism—a one-one correspondence that preserves mathematical structure—between any two models of the system. Thus if $\mathfrak{A}$ and $\mathfrak{B}$ are models of Q_E plus the principle of induction, let *f* be the smallest class that includes the pair $<0^{\mathfrak{A}}, 0^{\mathfrak{B}}>$ and includes $<S^{\mathfrak{A}}(x), S^{\mathfrak{B}}(y)>$ whenever it contains $<x,y>$. It is easy to verify, using induction several times, that *f* is an isomorphism. It follows that the system is complete, since if it left ϕ undecided, it would have a model $\mathfrak{A}$ in which ϕ is true and a model $\mathfrak{B}$ in which ϕ is false; but then $\mathfrak{A}$ and $\mathfrak{B}$ could not be isomorphic.

Peano Arithmetic (PA), is the system used to formalize the principle of induction into a precise system of axioms. Its axioms are Q_E together with all instances of the *induction axiom schema*:

$$((R(0) \wedge (\forall x)(R(x) \rightarrow R(Sx))) \rightarrow (\forall x)R(x)).$$

An *induction axiom* is a sentence of the language of arithmetic obtained from the schema by substituting a formula of the language of arithmetic for "*R*," then prefixing universal quantifiers to bind all the variables other than "*x*" that appear free in the substituted formula.

In view of Dedekind's categoricity theorem, it is surprising to realize that PA is incomplete. But incomplete it must be, since it is a true, recursively axiomatized theory that entails Q_E. The explanation is that the induction axiom schema does not fully capture the principle of

mathematical induction. It tries to assure us that every nonempty collection has a least element, but only succeeds in telling us that every nonempty collection that is the extension of a predicate (with parameters) of the language of arithmetic has a least element.

Let γ be the Gödel sentence for PA. We know that γ isn't a consequence of PA, so that, by the Completeness Theorem, there is a model $\mathfrak{A}$ in which all the axioms of PA + ~ γ are true. In $\mathfrak{A}$ there is an element *g* that satisfies "$x\mathrm{B}_{\mathrm{PA}}\ulcorner\underline{\gamma}\urcorner$." For each *n*, "$\sim\underline{n}\mathrm{B}_{\mathrm{PA}}\ulcorner\underline{\gamma}\urcorner$." is a theorem of PA, so *g* must be different from the referents of all the numerals $\underline{0}$, $\underline{1}$, $\underline{2}$, …. Instead, *g* is one of the *nonstandard numbers* that lie above all the standard numbers in the relation $\mathfrak{A}$ assigns to "≤."

It is worth emphasizing because there has been some confusion on this score that the existence of nonstandard models of PA does not depend on the First Incompleteness Theorem. Their existence follows from the Compactness Theorem, according to which an infinite set of sentences has a model if every finite subset does, which Gödel derived from the Completeness Theorem. Let Γ be a consistent theory that entails Q_{E}. Add a new constant "*c*" to the language, and let Γ^{c} be the union of Γ with the set of sentences "~ $c = \underline{n}$," for *n* natural number. Any finite subset of Γ^{c} has a model, obtained by taking a model of Γ and letting "*c*" denote a sufficiently large standard number. The Compactness Theorem gives us a model of Γ^{c}, which means we have a nonstandard model of Γ. This construction works even if we take Γ to be *true arithmetic*, the set of sentences true in the standard model, even though true arithmetic is complete. Because it is complete, the First Incompleteness Theorem tells us that true arithmetic is not recursively axiomatizable.

THE SECOND INCOMPLETENESS THEOREM

The proof of the First Incompleteness Theorem showed that, if Γ, a recursively axiomatized theory that entails Q_{E}, is consistent, then the Gödel sentence γ for Γ is unprovable in Γ. Using "*Con*(Γ)" as an abbreviation for

$$\text{“}\sim \mathit{Bew}_{\Gamma}(\ulcorner\underline{\underline{\sim \emptyset=\emptyset}}\urcorner)\text{,”}$$

we can formalize this result in a sentence of the language of arithmetic:

$$(\mathit{Con}(\Gamma) \rightarrow \sim \mathit{Bew}_{\Gamma}(\ulcorner\underline{\gamma}\urcorner)).$$

If we were able to prove this conditional in Γ, we could conclude that, if *Con*(Γ) were provable in Γ, ~ $\mathit{Bew}_{\Gamma}(\ulcorner\underline{\gamma}\urcorner)$ would be provable in Γ. Since we already know that ~ $\mathit{Bew}_{\Gamma}(\ulcorner\underline{\gamma}\urcorner)$ is only provable in Γ if Γ is inconsistent, we could conclude that *Con*(Γ) is only provable in Γ if Γ is inconsistent.

Can we prove the conditional in Γ? We certainly cannot do so if we take Γ to be Q_{E}, for we can scarcely prove any significant generalizations in Q_{E}. We can, however, prove the conditional if we take Γ to be PA. This is hardly surprising, since nearly all our reasoning about natural numbers can be formalized in PA. The details are, nonetheless, burdensome; so we only present a faint sketch here.

Let Γ be a recursively axiomatized theory that entails PA. M. H. Löb singled out the following three principles as central to Gödel's proof that, if Γ is consistent, it does not prove *Con*(Γ):

(L1) If $\Gamma \vdash \phi$, then $\Gamma \vdash \mathit{Bew}_{\Gamma}(\ulcorner\underline{\phi}\urcorner)$.

(L2) $\Gamma \vdash (\mathit{Bew}_{\Gamma}(\ulcorner\underline{\phi}\urcorner) \rightarrow \mathit{Bew}_{\Gamma}(\ulcorner\underline{\mathit{Bew}_{\Gamma}(\ulcorner\underline{\phi}\urcorner)}\urcorner))$.

(L3) $\Gamma \vdash (\mathit{Bew}_{\Gamma}(\ulcorner\underline{(\phi \rightarrow \Psi)}\urcorner) \rightarrow (\mathit{Bew}_{\Gamma}(\ulcorner\underline{\phi}\urcorner) \rightarrow \mathit{Bew}_{\Gamma}(\ulcorner\underline{\Psi}\urcorner)))$.

We have already seen why (L1) has to hold. If φ is a consequence of Γ, $\mathit{Bew}_{\Gamma}(\ulcorner\underline{\phi}\urcorner)$ is a true Σ sentence, hence provable in Q_{E}, hence provable in Γ. (L2) is obtained, laboriously, by formalizing the proof of (L1). In fact, Γ proves $(\theta \rightarrow \mathit{Bew}_{\Gamma}(\ulcorner\underline{\theta}\urcorner))$, for each Σ sentence θ. (L3) is easy. If we have proofs of $(\phi \rightarrow \psi)$ and φ, we get a proof of ψ by concatenating the two proofs and tacking ψ on the end.

Given the Löb conditions, the proof of the *Second Incompleteness Theorem*, according to which, if Γ is a consistent, recursively axiomatized theory that entails PA, then Γ does not prove its own consistency, is straightforward. Let γ be the Gödel sentence for Γ. Because of the way γ was constructed, we have:

$$\Gamma \vdash (\gamma \rightarrow \sim \mathit{Bew}_{\Gamma}(\ulcorner\underline{\gamma}\urcorner)),$$

which is logically equivalent to:

$$\Gamma \vdash (\gamma \rightarrow (\mathit{Bew}_{\Gamma}(\ulcorner\underline{\gamma}\urcorner) \rightarrow \sim 0=0)).$$

One application of (L1) and two applications of (L3) give us this:

$$\Gamma \vdash (\mathit{Bew}_{\Gamma}(\ulcorner\underline{\gamma}\urcorner) \rightarrow (\mathit{Bew}_{\Gamma}(\ulcorner\underline{\mathit{Bew}_{\Gamma}(\ulcorner\underline{\gamma}\urcorner)}\urcorner) \rightarrow \mathit{Bew}_{\Gamma}(\ulcorner\underline{\sim \emptyset=\emptyset}\urcorner))).$$

(L2) gives us this:

$$\Gamma \vdash (\mathit{Bew}_{\Gamma}(\ulcorner\underline{\gamma}\urcorner) \rightarrow \mathit{Bew}_{\Gamma}(\ulcorner\underline{\mathit{Bew}_{\Gamma}(\ulcorner\underline{\gamma}\urcorner)}\urcorner)),$$

and these two results together give us:

$$\Gamma \vdash (\mathit{Bew}_{\Gamma}(\ulcorner\underline{\gamma}\urcorner) \rightarrow \mathit{Bew}_{\Gamma}(\ulcorner\underline{\sim \emptyset=\emptyset}\urcorner)).$$

By contraposition,

$$\Gamma \vdash (\sim \mathit{Bew}_{\Gamma}\ulcorner\underline{\sim \emptyset=\emptyset}\urcorner) \rightarrow \sim \mathit{Bew}_{\Gamma}(\ulcorner\underline{\gamma}\urcorner)),$$

that is,

$$\Gamma \vdash (\mathit{Con}(\Gamma) \rightarrow \sim \mathit{Bew}_{\Gamma}(\ulcorner\underline{\gamma}\urcorner))$$

Now assume

$$\Gamma \vdash Con(\Gamma).$$

Then

$$\Gamma \vdash \sim Bew_{\Gamma}(\ulcorner \underline{\gamma} \urcorner).$$

By the way γ was constructed,

$$\Gamma \vdash \gamma.$$

Hence, by (L1),

$$\Gamma \vdash Bew_{\Gamma}(\ulcorner \underline{\gamma} \urcorner),$$

and so Γ is inconsistent.

In accepting PA, we recognize that the axioms of PA are all true. If the axioms are all true then the theory is certainly consistent, and if the theory is consistent its Gödel sentence is true. So we have good reason to accept the Gödel sentence for PA, even though it is not a consequence of PA. If in this argument we replace PA with our total arithmetical theory—the (admittedly, vaguely defined) totality of arithmetical sentences we are willing to accept as true—we seem to get the curious result that, assuming that our total theory is recursively enumerable, we accept the Gödel sentence for our total theory even though it is not a consequence of the theory. But this contradicts the characterization of our total theory.

J. R. Lucas (1961) and Roger Penrose (1989) took this puzzling situation as reason to believe that the cognitive processes of the human mind cannot be simulated by any purely mechanical device, and that this conclusion undermines the prospects for a naturalistic conception of mind, according to which the human mind is a product of the orderly operation of the laws of nature, not in principle any more mysterious or less constrained by physical law than a player piano or a personal computer. Adherents to the computational theory of mind hold that the operations of the mind are usefully understood on the model of a sophisticated electronic computer, and even naturalists who are not advocates of the computational model will be inclined to say that the facts that the human body is produced by natural selection rather than conscious design and that its central processing unit is carbon-based rather than silicon-based will not affect its capabilities in any fundamental way, so that, according to a naturalistic conception, the cognitive activities of a human being can, in principle, be simulated by a purely mechanical device.

The connection between mechanism and recursive enumerability is given by a variant of the Church-Turing Thesis, supported by similar evidence, that declares that the set of numbers accepted by a mechanical input-output device is invariably recursively enumerable. This includes nondeterministic machines, whose operation is to some extent a matter of random chance, so that the set *S* is accepted by the machine just in case, for any *n*, *n* is in *S* if and only if there is some possible computation of the machine on input *n* that yields a positive outcome, as well as deterministic machines for which the course of a computation is uniquely determined by its input.

The argument that our total arithmetical theory is not recursively enumerable proceeds by *reductio ad absurdum*. If the theory were recursively enumerable, it would be recursively axiomatizable, so it would have a Gödel sentence. But we can see that the Gödel sentence is true, even though it is not part of the total theory.

The Lucas-Penrose argument is vulnerable to two criticisms. First, for naturalism to be correct, there has to exist a recursive axiomatization of our total theory. In order to construct the Gödel sentence, we have to be able to specify a recursive axiomatization by writing down a formula that binumerates it. However it is perfectly possible for a recursive axiomatization to exist without our being able to specify it.

Second, even if we were able to specify a recursive axiomatization, perhaps by analyzing a futuristic brain scan, it is hard to see how we could be justified in being completely confident that our total theory is consistent. If we decide to be strict about what arithmetical sentences we are willing to count as "accepted," so that we only regard a sentence as part of our total theory if we arrive at it by unimpeachably lucid reasoning, we shall increase our confidence that our total theory is consistent, but raising the bar this way will also heighten the hurdle that the Gödel sentence has to pass in order to count as "accepted." There are different standards we might use for when we are willing to count a sentence as proven, and each standard has a different Gödel sentence, but however high we set the standard the Gödel sentence corresponding to that standard cannot pass it, on pain of inconsistency.

THE LOGIC OF PROVABILTY

If we explicitly embrace a theory Γ, so that we are willing consciously to acknowledge that the axioms of Γ are all statements we regard as true then we surely ought to regard Γ as consistent. Yet (assuming that Γ implies PA and is recursively axiomatizable and consistent) the statement that Γ is consistent is not provable in Γ. Thus the arithmetical statements that we commit ourselves to in embracing Γ go beyond what Γ itself entails.

The disparity between what consciously accepting Γ commits us to and what Γ entails is even wider than the Second Incompleteness Theorem indicates. Accepting Γ means acknowledging that all the consequences of Γ are true. For a given sentence ϕ, we may not know whether ϕ is a consequence of Γ—there is after all no algorithm to tell us—but at least we accept that, if ϕ is a consequence of Γ, ϕ is true. Consciously accepting Γ commits us to the conditionals $Bew_\Gamma(\ulcorner\underline{\phi}\urcorner) \rightarrow \phi)$, but they are not in general consequences of Γ. In fact such a conditional is a consequence of Γ only if its consequent is a consequence of Γ.

Löb's Theorem. Let Γ be a recursively axiomatized theory that entails PA. If $Bew_\Gamma(\ulcorner\underline{\phi}\urcorner) \rightarrow \phi)$ is a consequence of Γ, so is ϕ.

We can regard the Second Incompleteness Theorem as the special case of Löb's Theorem in which ϕ is taken to be the sentence "~ 0=0." Conversely we can derive Löb's Theorem from the Second Incompleteness Theorem. The argument, which is due to Saul Kripke, utilizes the observation that, for any ψ and θ, $\Gamma \vdash (\psi \rightarrow \theta)$ if and only if $\Gamma \cup \{\psi\} \vdash \theta$, and the fact that this observation is provable in PA.

Suppose that ϕ is not a consequence of Γ. Then $\Gamma \cup \{\sim \phi\}$ is consistent, which implies, by the Second Incompleteness Theorem, that $Con(\Gamma \cup \{\sim \phi\})$ is not a consequence of $\Gamma \cup \{\sim \phi\}$. Thus we have:

$\Gamma \cup \{\sim \phi\} \nvdash \sim Bew_{\Gamma \cup \{\sim\phi\}}(\ulcorner\underline{\sim 0=0}\urcorner)$

$\Gamma \cup \{\sim \phi\} \nvdash \sim Bew_{\Gamma}(\ulcorner\underline{(\sim \phi \rightarrow \sim 0=0}\urcorner)$

$\Gamma \cup \{\sim \phi\} \nvdash \sim Bew_{\Gamma}(\ulcorner\underline{\phi}\urcorner)$

$\Gamma \nvdash (\sim \phi \rightarrow \sim Bew_{\Gamma}(\ulcorner\underline{\phi}\urcorner))$

$\Gamma \nvdash Bew_{\Gamma}(\ulcorner\underline{\phi}\urcorner) \rightarrow \phi)$

Conditionals of the form $Bew_\Gamma(\ulcorner\underline{\phi}\urcorner) \rightarrow \phi)$ are called *reflection principles*. We cannot obtain them by working within Γ. We get them from the outside by reflecting on the fact that Γ is a theory we accept.

We can describe the logic of provability precisely by utilizing the methods of modal logic. Modal sentential calculus has, in addition to formulas built up from atomic formulas by the familiar connectives "∨" and "~," a new connective "□." "□ϕ," usually read "It is necessary that ϕ," is here understood to mean, "It is provable in Γ that ϕ," where Γ is a consistent, recursively axiomatizable theory that implies PA. An *interpretation* of the modal sentential calculus is a function *i* that associates an arithmetical sentence with each modal formula, subject to the conditions that $i(\phi \vee \psi)$ be equal to $(i(\phi) \vee i(\psi))$, $i(\sim\phi)$ be equal to $\sim i(\phi)$, and $i(\Box\phi)$ be equal to $Bew_\Gamma\ulcorner\underline{i(\phi)}\urcorner$. A modal formula ϕ is *always provable* if, for each interpretation *i*, $i(\phi)$ is provable in Γ. ϕ is *always true* if, for each ϕ, $i(\phi)$ is true.

(L1) tells us, if $i(P)$ is provable, $i(\Box P)$ is provable, so that the set of always-provable formulas is closed under necessitation, the rule of modal logic that infers □θ from θ. (L2) tells us that $(\Box P \rightarrow \Box\Box P)$ is always true, and the formalization of (L2) tells us that it is always provable. (L3) tells us that $(\Box(P \rightarrow Q) \rightarrow (\Box P \rightarrow \Box Q))$ is always true; it is easily seen to be always provable as well. Löb's Theorem tells us that whenever $i(\Box P \rightarrow P)$ is a theorem, $i(P)$ is a theorem. Formalizing his proof, we see that the formula $(\Box(\Box P \rightarrow P) \rightarrow \Box P)$ is always provable and always true.

Robert Solovay deployed an ingenious application of the Self-referential Lemma within the possible-world semantics for modal logic to show that, provided Γ does not prove any false Σ sentences, a formula is always provable if and only if it is derivable by *modus ponens* and necessitation from sentential-calculus tautologies (formulas that are assigned the value "true" by every function assigning truth-values to formulas that respects the meanings of "∨" and "~") and instances of the following schemata:

$(\Box(\phi \rightarrow \psi) \rightarrow (\Box\phi \rightarrow \Box\psi))$

$(\Box(\Box\phi \rightarrow \phi) \rightarrow \Box\phi)$

Assuming Γ is true, a formula is always true if and only if it is derivable by *modus ponens* from always-provable formulas and instances of the reflection principle $(\Box\phi \rightarrow \phi)$.

BEYOND THE LANGUAGE OF ARITHMETIC

Gödel's results apply not only to the language of arithmetic but to any language into which the language of arithmetic can be translated. Thus any recursively axiomatized, consistent theory into which one can translate Q_E is incomplete. The appropriate notion of translation was made precise by Tarski, Mostowski, and Robinson. An *interpretation* (what they call a "relative interpretation") of an arithmetical theory Γ into a language $\mathfrak{L}$ is obtained by doing the following: First, having rewritten all the sentences in Γ so that the "+" sign only appears in the canonical form "$(v_i + v_j) = v_k$," pick a formula "$A(x,y,z)$" of $\mathfrak{L}$ and replace "$(v_i + v_j) = v_k$," by "$A(v_i,v_j,v_k)$," changing bound variables to avoid conflicts. Do the same thing for the other function signs and "0" and pick a formula $L(x,y)$ to replace "≤." Next pick a formula "$N(x)$" of $\mathfrak{L}$ to represent the members of the domain of $\mathfrak{L}$ that are to play the role of natural numbers, and restrict the quantifiers, writing "$(\exists v_i)(N(v_i \wedge \ldots$" in place of $(\exists v_i)$. Finally add an axiom ensuring that "$A(x,y,z)$" represents a function on

the set of things that satisfy "$N(x)$," writing "$(\forall x)(N(x) \rightarrow (\forall y)(\mathrm{N}(y) \rightarrow (\exists z)(N(z) \wedge (\forall w)(N(w) \rightarrow (A(x,y,w) \leftrightarrow w = z))))$." Do the same thing for the other function signs and "0." If the theory thus obtained is a consequence of the theory Δ of $\mathfrak{L}$, Δ is said to *interpret* Γ.

We can translate the language of arithmetic into the language of set theory, identifying a number with the set of its predecessors, so that 0 corresponds to $\emptyset$, 1 corresponds to $\{\emptyset\}$, 2 corresponds to $\{\emptyset, \{\emptyset\}\}$, and so on, and defining set-theoretic analogues of "+," "×," "E," "*S*," and "≤" accordingly. The axioms of set theory, in any of its normal versions, interpret PA. We can arithmetize proofs in set theory just as we artimeticized proofs in PA, proving the Second Incompleteness Theorem for set theory. The axioms of set theory, if consistent, cannot prove their own consistency.

This result devastates the Hilbert program. Hilbert wanted to prove the consistency of set theory in a finitistic theory much weaker than set theory, and it turns out that proving the consistency of set theory requires a theory even stronger than set theory.

The standard way to prove that there is no algorithm for testing whether a given sentence is a consequence of a theory Γ—that is, for showing that Γ is *undecidable*—is to interpret an arithmetical theory strong enough to prove the First Incompleteness Theorem into Γ. As far as what we have looked at so far, we would need to take our arithmetical theory to be $\mathrm{Q_E}$, but we can actually do much better. We can define exponentiation in terms of "0," "*S*," "+," and "×," and we can prove the First Incompleteness Theorem in the dialect of the language of arithmetic without "E," with Q in place of $\mathrm{Q_E}$. In trying to prove undecidablity results, this improvement (which is due to Gödel) is an enormous practical advantage.

Let us define $\beta(u,v,w)$ to be the remainder obtained on dividing u by $(v \times w) + 1$. β can be defined by a bounded formula in the language of arithmetic. For $x > 0$, we have $(x\mathrm{E}y) = z$ if and only if the following formula is satisfied:

$$(\exists u)(\exists v)((\beta(u,v,0) = 1 \wedge (\forall w < y)\beta(u,v,\mathrm{S}w) = (\beta(u,v,w) \times x)) \wedge \beta(u,v,y) = z).$$

The right-to-left direction of this characterization is obvious. What is hard is to verify the left-to-right direction by finding an appropriate u and v. We make use of the Chinese Remainder Theorem, which says that, given $p_0, p_1, \ldots, p_n$ relatively prime (that is, no two of the p_is have a common divisor other than 1), and given a sequence $a_0, a_1, \ldots, a_n$, with each $a_i < p_i$, we can find a number b such that a_i is the remainder on dividing b by p_i, for each i. A proof of the theorem can be found in any number-theory textbook or in George Boolos's *The Logic of Provability* (1993).

Given x,y, and z with $(x\mathrm{E}y) = z$, let $v = z!$, the product of the positive integers $\leq z$. If $s < t \leq z$, then $(s \times v) + 1$ and $(t \times v) + 1$ are relatively prime, since if p were a prime that divided both of them, p would divide $(t - s) \times v$, and so, since $(t - s)$ is one of the factors of v, p would divide v. But this enables us to conclude that the remainder on dividing $(t \times v) + 1$ by p is one, contrary to our assumption that p divides $(t \times v) + 1$. Use the Chinese Remainder Theorem to find u so that, for each $t \leq y$, $x\mathrm{E}t$ is the remainder on dividing u by $(t \times v) + 1$.

Now that we have our Σ definition of exponentiation—Σ, that is, in the restricted language—we can apply our standard tricks for pulling quantifiers to the fronts of formulas to convert a Σ formula of the language with exponentiation to a Σ formula of the language without exponentiation. With this emendation, all the proofs go through.

The use of interpretations originates with Beltrami's proof of the consistency of non-Euclidean geometry. By interpreting non-Euclidean geometry (Euclid's axioms with the axiom of parallels replaced by its negation) into Euclidean geometry, Beltrami showed that if the latter is consistent then so is the the former. Beltrami's strategy was exploited by Alex Wilkie and Samuel Buss to obtain a dramatic strengthening of the Second Incompleteness Theorem, applying it to theories that merely contain Q rather than PA. The details are complicated, but the idea is to interpret into Q a theory that, while weaker than PA (the induction axiom schema being restricted), is just strong enough to provide the Löb conditions (L1)-(L3). The interpretation leaves the arithmetical symbols unchanged but restricts the domain of quantification to an initial segment, replacing "$(\exists x)$" by "$(\exists x)(J(x) \wedge \ldots$," for artfully chosen "$J(x)$;" call the sentence thus obtained from ϕ "ϕ^J."

Where Γ is a recursively axiomatized theory that includes Q, let Γ^J be the set of sentences ϕ for which Γ entails ϕ^J. Suppose that Γ entails $\mathrm{Con}(\Gamma)$. $\mathrm{Con}(\Gamma)$ entails $\mathrm{Con}(\Gamma)^J$, so that $\mathrm{Con}(\Gamma)$ is in Γ^J. The argument Beltrami used tells us that if Γ is consistent then Γ^J is too. This proof can be formalized in Γ^J, so that Γ^J entails $\mathrm{Con}(\Gamma^J)$. Because (L1)-(L3) yield the Second Incompleteness Theorem for Γ^J, Γ^J must be inconsistent. Consequently Γ is inconsistent.

TRUTH

There is a bounded formula of the language of arithmetic that defines the set of prime numbers, and there is a Σ for-

mula that defines the set of consequences of PA. Tarski proved that there is no formula of the language of arithmetic that defines the set of codes of true sentences. The difficult part of his argument was to say precisely what would be required for a formula to define truth; the easy part is to show that there is no such formula.

A proposed definition of truth is a formula of the form $(Tr(x) \leftrightarrow \tau(x))$, where $\tau(x)$ is a formula of the language of arithmetic. A proposed definition is *materially adequate*, Tarski tells us, if and only if it lets us derive all sentences of the form:

(T) $$Tr(\ulcorner \phi \urcorner) \leftrightarrow \phi.$$

To see that there is no materially adequate definition, apply the Self-reference Lemma to find a sentence λ so that $(\lambda \leftrightarrow \sim\tau(\ulcorner \underline{\lambda} \urcorner))$ is a consequence of Q. The argument here is a formalization of the paradox posed by Eubulides, who asked whether a man who says "I am lying" speaks truthfully.

We can define the set of true sentences of the language of arithmetic within, say, the language of set theory, but we cannot define it within the language of arithmetic. This negative result obtains for any language into which we can translate the language of arithmetic.

The question of what moral, if any, these formal results have for the notion of truth as applied to natural languages is deeply troubling. Tarski showed that there is no formula of the language of arithmetic that means (or even has the same extension as) "true sentence of the language of arithmetic." Manifestly there is a phrase of English that means "true sentence of English," and Tarski and Eubulides' reasoning would appear to apply to that phrase just as to the formal language. Is there in spite of this a coherent way to talk about the truth of an English sentence?

See also Analysis, Philosophical; Aristotle; Church, Alonzo; Computability Theory; Craig's Theorem; Geometry; Gödel, Kurt; Hilbert, David; Infinity in Mathematics and Logic; Kripke, Saul; Logic, History of: Modern Logic; Logical Paradoxes; Mathematics, Foundations of; Russell, Bertrand Arthur William; Tarski, Alfred; Turing, Alan M.; Wittgenstein, Ludwig Josef Johann; Zeno of Elea.

Bibliography

Aristotle. *Prior Analytics*. In *Complete Works of Aristotle*, edited by Jonathan Barnes. Princeton, NJ: Princeton University Press, 1995.

Beltrami, Eugenio. "Saggio di Interpretazione della Geometria Noneuclidea." *Giornale di Matematiche* 6 (1868): 284–312.

Boolos, George *S. The Logic of Provability*. Cambridge, U.K.: Cambridge University Press, 1993.

Buss, Samuel R. "First-Order Proof Theory of Arithmetic." In *Handbook of Proof Theory*, edited by Samuel R. Buss, 79–147. Amsterdam-Elsevier, 1998.

Church, Alonzo. "A Note on the Entscheidungsproblem." *Journal of Symbolic Logic* 1 (1936): 40–41, with a correction on pp. 101–102. Reprinted in *The Undecidable,* edited by Martin Davis, 110–115. Hewlett, NY: Raven Press, 1965.

Craig, William. "On Axiomatizability Within a System." *Journal of Symbolic Logic* 18 (1953): 30–32.

Davis, Martin. *The Undecidable*. Hewlett, NY: Raven Press, 1965.

Dedekind, Richard. *Was sind und was sollen die Zahlen?* Brunswick: Vieweg, 1888. Reprinted in Dedekind, *Essays on the Theory of Numbers*. New York: Dover, 1963, 29–115.

Euclid. *Elements*. 2 vols. Translated by Thomas L. Heath. 2nd ed. New York: Dover, 1956.

Frege, Gottlob. *Begriffschrift*. Halle: L. Nebert, 1879. Reprinted in *From Frege to Gödel*, edited by Jean van Heijenoort. 1–82. Cambridge, MA: Harvard University Press, 1967.

Gödel, Kurt. "Die Vollständigkeit der Axiome der logischen Funktionenkalküls." *Monatschefte für Mathematik un Physik* 37 (1930): 349–360. Reprinted in *From Frege to Gödel,* edited by Jean van Heijenoort (1967), 583–591. Also reprinted in Gödel, *Collected Works,* vol. I, 144–195. New York: Oxford University Press, 1986.

Gödel, Kurt. "Über formal unentscheidbare Sätze der *Principia mathematica und verwandter Systeme* I." *Monatshefte für Mathematik und Physik* 38 (1931): 173–198. Reprinted in *The Undecidable*, edited by Martin Davis, 4–38. Hewlett, NY: Raven Press, 1965. Also reprinted in *From Frege to Gödel,* edited by Jean van Heijenoort (1967), 596–628. Also reprinted in Gödel, *Collected Works,* vol. I, 102–123. New York: Oxford University Press, 1986.

Hájek, Petr, and Pavel Pudlák. *Metamathematics of First-Order Arithmetic*. New York: Springer, 1993.

Hilbert, David. "Über dan Undendliche." *Mathematische Annalen* 95 (1926): 161–190. Reprinted in *From Frege to Gödel*, edited by Jean van Heijenoort. 367–392. Cambridge, MA: Harvard University Press, 1967.

Löb, M. H. "Solution to a Problem of Leon Henkin." *Journal of Symbolic Logic* 20 (1955): 115–118.

Lucas, J. R. "Minds, Machines, and Gödel." *Philosophy* 36 (1961): 120–124.

Penrose, Roger. *The Emperor's New Mind*. New York: Oxford University Press, 1989.

Quine, Willard van Orman. *Mathematical Logic*. 2nd ed. Cambridge, MA: Harvard University Press, 1951.

Rosser, John Barclay. "Extensions of Some Theorems of Gödel and Church." *Journal of Symbolic Logic* 1 (1936): 87–91. Reprinted in *The Undecidable*, edited by Martin Davis, 231–235. Hewlett, NY: Raven Press, 1965.

Smorynski, Craig. "The Incompleteness Theorems." In *Handbook of Mathematical Logic*, edited by Jon Barwise, 821–865. New York: North-Holland, 1977.

Smullyan, Raymond M. *Gödel's Incompleteness Theorems*. New York: Oxford University Press, 1992.

Solovay, Robert M. "Provability Interpretations of Modal Logic." *Israel Journal of Mathematics* 25 (1976): 287–304.

Tarski, Alfred. "Der Wahrheitsbegriff in den formalisierten Sprachen." *Studia Philosophica* 1 (1935): 261–405. Reprinted

in Tarski, *Logic, Semantics, Metamathematics*, 2nd ed., 152–278. Indianapolis: Hackett, 1983.

Tarski, Alfred, Andrzej Mostowski, and Raphael M. Robinson. *Undecidable Theories*. Amsterdam: North-Holland, 1953.

Turing, Alan Mathison. "On Computable Numbers, with an Application to the Entscheidungsproblem." *Proceedings of the London Mathematical Society*, 2nd series, 42 (1937): 230–265, with a correction at vol. 43 (1938): 544–546. Reprinted in *The Undecidable*, edited by Martin Davis, 115–154. Hewlett, NY: Raven Press, 1965.

Van Heijenoort, Jean. *From Frege to Gödel*. Cambridge, MA: Harvard University Press, 1967.

Wilkie, Alex J., and Jeff B. Paris. "On the Scheme of Induction for Bounded Arithmetic Formulas." *Annals of Pure and Applied Logic* 35 (1987): 261–302.

Vann McGee (2005)

GODFREY OF FONTAINES

Godfrey of Fontaines, the scholastic philosopher and theologian, was a native of Fontaines-les-Hozémont in the principality of Liège. He was born of a noble family about the middle of the thirteenth century, the exact date unknown. About 1270 he began studies at the University of Paris and became a *magister regens* in the faculty of theology there in 1285, having studied under Henry of Ghent and Gervais of Mt. St. Elias. His regency lasted until 1297, and during this period he produced fourteen of his *Quodlibets,* his most important works. There is evidence that he resumed teaching at Paris about 1303 or 1304, composing *Quodlibet* XV at this time. Canon of Liège, probably also of Paris, and provost of Cologne (1287–1298), Godfrey was chosen bishop of Tournai in 1300 but renounced his rights when the election was contested. He is cited among the senior members of the Sorbonne until 1306 and probably died about that time. The obituary at the Sorbonne dates his death October 29, but does not give the year.

Godfrey's doctrinal preferences generally favor the positions of St. Thomas Aquinas, but he manifests a marked independence of judgment on certain points and sometimes works out the logic of Thomas's principles to different conclusions. Some historians (M. De Wulf, E. Gilson) see Godfrey as an opponent of Thomas's distinction between essence and existence in finite being, and attribute Godfrey's stand to a hard-and-fast Aristotelianism that refused to admit an act of the form. Others see Godfrey as opposing the realism of Giles of Rome rather than Thomas. Godfrey held that in the divine mind there is no proper idea of individuals distinct from their species. On the hotly debated issue of the oneness or plurality of substantial forms in composite beings, Godfrey always remained hesitant. He would have favored the doctrine of the unicity of form were it not for the fact that it seemed to contradict theological truths.

Godfrey showed particular acumen in his treatment of psychological problems. Under the influence of Averroes, probably through Siger of Brabant, he espoused an Aristotelianism stricter than that of most of his contemporaries. Godfrey criticized and rejected the so-called Augustinian theory on the genesis of ideas, insisting on the close dependence of human concepts on sense experience. He insisted strongly on the passive nature of the human intellect—the abstractive function of the agent intellect does not consist in the production of any positive disposition in the sensible image upon which it works, but in disregarding in a merely negative way the concrete particularizations characteristic of the image. This outlook is intimately connected with an Avicennan realism of abstract essence, so that Godfrey held that the intellect does not produce intelligibility or universality either in things or in images, but that the agent intellect places the images under an illumination such that the quiddity or essence of the object can appear alone and act on the possible intellect and become known to us.

In his explanation of human free will Godfrey adhered closely to the Thomistic doctrine, but he insisted more than Thomas upon the freedom of the intellect as its foundation. Against the voluntarism of Henry of Ghent, Godfrey stressed the formal influence of the intellect upon the will to the point of making it an efficient cause, whereas Thomas, in different historical circumstances against the Averroists, minimized the formal influence of the object upon the will. In other respects Godfrey did not break cleanly with the Augustinian tradition. For example, he made an interesting equivalence of the active and passive intellects with Augustine's "memory," the passive intellect inasmuch as it conserves species and is a *habitus,* the active intellect inasmuch as it contributes to actual knowledge.

Godfrey was a lively controversialist, combating at length the opinions of his contemporaries, particularly Henry of Ghent, Giles of Rome, and James of Viterbo. Not only did he engage in an active dialogue with his contemporaries, but he also occupied himself with pressing problems—moral, legal, social, and political—arising from daily life. Among his admirers can be listed John the Wise, Peter of Auvergne, and Gerard of Bologna; among his critics, Bernard of Auvergne, Gonsalvus of Spain, and John Duns Scotus. His influence was widespread and lasted well into the fourteenth century but waned thereafter.